Supplementary Materials S1
A PDF file containing:
Part A: Additional Information on MU-Referencing
Part B: No-U-Turn Hamiltonian Bayes Method
Part C: Providing Prior Information for Bayesian Analysis
Part D: Termination Testing
Part E: Setting up Example 1 for Bayesian Analysis (504b.ctl)

Part A: Additional Information on MU-Referencing

Additional rules for applying MU referencing are as follows:

Rule Violation:

```
MU_1=THETA(1)
MU_2=THETA(2)
MU_3=THETA(3)
CL=EXP(MU_1+ETA(2))
V=EXP(MU_2+MU_3+ETA(1))
```

MU_1 is used as a mean to ETA(2)

Composite MU_2 and MU_3 are means to ETA(1)

Theta that models a MU_ may not show up in any subsequent lines of code. For example, if

```
CL=EXP(THETA(5)+ETA(2))
```

So that it can be rephrased as

```
MU_2=THETA(5)
CL=EXP(MU_2+ETA(2))
```

Later, THETA(5) is used without its association with ETA(2) (not correct):

...

```
CLZ=EXP(THETA(5))*2
```

But this is correct:

```
CLZ=CL*2
```

Parameter has fixed effect with no random effect:

Consider fixed effect parameters that have no etas associated with them, such as

```
Km=THETA(5)
```

The THETA(5) and Km cannot be MU referenced. In such cases, NONMEM will use a slower gradient method to move the theta towards maximization of the objective function. Mu referencing THETA(5) and setting its OMEGA to 0 will not help, as EM maximization requires random sets of individual parameters with some variability to advance THETA(5). Alternatively, an annealing technique may be implemented in NONMEM ($ANNEAL, intro7.pdf, guide VIII [1]). Generally, EM methods are still able to perform reasonably efficiently if there are no more than 3 thetas that are not MU-referenced. The FOCE tends to be more efficient in evaluating problems that have more than 3 thetas that are not mu-referenced.

Parameters with Covariates that vary within a Subject

Suppose the covariate WT changes with each observation record in a subject, and is used in a MU-reference equation.

```
MU_3=THETA(1)+LOG(WT)
```

There can only be one typical value, or MU relation, for a subject, so NONMEM will obtain a MU value from the average of the records for that subject. If such an approximation process is not acceptable, move the within-subject dependent (or time dependent) covariate from the typical value level to individual value level:

```
MU_3=THETA(1)
V=WT*EXP(MU_3+ETA(3))
```

In the above example, MU_3 now represents volume per unit weight, which is constant for the individual, while volume itself varies in proportion to weight, which may fluctuate during the study period.

Part B: No-U-Turn Hamiltonian Bayes Method

The No U-Turn sampling method of Bayesian Analysis [7,8] is useful when there are high correlations in the population parameters, which could be due to high correlations in the OMEGA block. The NUTS sampler uses a directed search using Partial derivatives and scaling techniques using posterior density knowledge from previous samples to reduce the correlation of the parameters from one iteration to the next. While each iteration takes longer to generate with NUTS, the samples may be several times decorrelated relative to a standard MCMC sampling. It is important to MU reference all thetas, so that NONMEM can evaluate their derivatives analytically, otherwise considerable computational inefficiency will result. Unlike Gibbs sampling method, the MU referencing need not be linear with respect to the thetas to be efficient, and the associated OMEGA value can be zero valued.

Part C: Providing Prior Information for Bayesian Analysis

In NONMEM, the $PRIOR record is used to declare prior information is to be used.

$PRIOR NWPRI
This specifies the Normal-Wishart prior type
Priors to THETAS are assumed multivariate normal
Input Theta Prior means, and Theta Prior Variance-Covariance
Priors to OMEGAS are assumed Wishart distributed
Input Omega Prior means, and Omega Prior degrees of freedom
Priors to SIGMAS are assumed Wishart distributed
Input Sigma Prior means, and Sigma Prior degrees of freedom

Uninformative Priors

An example of providing uninformative priors is as follows:

Prior Values to Theta:

```
$THETAP (2 FIXED)x4
;Variance to Theta Prior Values (large variance means uninformative):
$THETAPV BLOCK(4)
```

```
10000 FIXED
0.0      10000
0.0         0.0    10000
0.0         0.0      0.0     10000
```

Or, use short hand:

```
$THETAPV BLOCK(4) FIXED VALUES(10000,0)
```

Prior information to the OMEGAS:

```
$OMEGAP BLOCK(4) FIXED VALUES(0.2,0.0)
```

Note that all inverse Wishart priors are at least weakly informative, and the values could be set to approximately that obtained from a previous maximum likelihood analysis result.

Degrees of Freedom (equals dimension of OMEGA block for uninformative):

```
$OMEGAPD (4.0 FIXED)
```

Prior information to the SIGMAS:

```
$SIGMAP 0.06 FIXED
```

Degrees of Freedom (equals dimension of OMEGA block for uninformative):

```
$SIGMAPD (1.0 FIXED)
```

Informative THETA Priors

Informative THETA priors may be obtained from THETA estimates from a previous study. From the report file of previous study:

```
 FINAL PARAMETER ESTIMATE
 THETA - VECTOR OF FIXED EFFECTS PARAMETERS
         TH 1      TH 2      TH 3      TH 4
         1.64E+00  1.57E+00  7.58E-01  2.35E+00
```

Place in the control stream for new study as follows:

```
$THETAP   (1.64 FIXED) (1.57 FIXED) (0.758 FIXED)
          (2.35 FIXED)
```

Theta Prior variance should be THETAxTHETA portion of variance-covariance of estimates from previous study. From report file of previous study:

```
COVARIANCE MATRIX OF ESTIMATE
            TH 1      TH 2      TH 3      TH 4
 TH 1
+        2.33E-03
 TH 2
+        4.76E-04  2.86E-03
 TH 3
+        7.87E-04  1.27E-04  5.35E-03
 TH 4
+        7.80E-05  2.36E-04  1.76E-03  2.98E-03
```

Place in control stream file of new study as follows:

```
$THETAPV BLOCK(4)
        2.33E-03 FIXED
        4.76E-04  2.86E-03
        7.87E-04  1.27E-04  5.35E-03
        7.80E-05  2.36E-04  1.76E-03  2.98E-03
```

OMEGA Prior should be Estimates of OMEGAS in previous study from report file:

```
 OMEGA - COV MATRIX FOR RANDOM EFFECTS - ETAS
            ETA1      ETA2      ETA3      ETA4
 ETA1
+        1.75E-01
 ETA2
+        8.33E-03  1.51E-01
 ETA3
+        2.98E-02  1.74E-02  2.41E-01
  ETA4
+       -8.05E-03  1.84E-02  5.14E-02  1.62E-01
```

Place in control stream file of new study as follows:

```
 $OMEGAP BLOCK(4)
    1.75E-01 FIXED
    8.33E-03  1.51E-01
    2.98E-02  1.74E-02  2.41E-01
   -8.05E-03  1.84E-02  5.14E-02  1.62E-01
```

OMEGA Prior Degrees of freedom should be no larger than total number of subjects in previous study
More accurately, Use the following relation: (courtesy of Mats Karlsson):
DF=2*[(OMEGA estimate of previous analysis)/(SE of OMEGA of previous analysis)]^2
Or,
DF=2*[(OMEGA estimate of previous analysis)/(SE of OMEGA of previous analysis)]^2+1
to adjust for degrees of freedom loss in the estimate of OMEGA of the previous study.
For an OMEGA block, use the smallest DF calculated among the OMEGA diagonal estimates in that block.
SIGMA Prior should be Estimates of SIGMAS in previous study from report file:

```
 SIGMA - COV MATRIX FOR RANDOM EFFECTS - EPSILONS  ***
            EPS1
 EPS1
+        5.28E-02
```

Place in control stream file of new study as follows:

```
 $SIGMAP (5.28E-02 FIXED)
```

SIGMA Prior Degrees of freedom should be no larger than total number of data points in previous study pertaining to that SIGMA, for example:
For PK SIGMAS, no. of PK data points
For PD SIGMAS, no. of PD data points
More accurately, Use the following relation:
DF=2*[(SIGMA estimate of previous analysis)/(SE of SIGMA of previous analysis)]^2
Or,
DF=2*[(SIGMA estimate of previous analysis)/(SE of SIGMA of previous analysis)]^2+1
to adjust for degrees of freedom loss in the estimate of Omega of the previous study.
For a SIGMA block, use the smallest DF calculated among the SIGMA diagonal estimates in that block.

Part D: Termination Testing

At each iteration, the program performs a linear regression on each parameter (x=iteration number, y=parameter value).
If the slope of the linear regression is not statistically different from 0 for all parameters tested, then convergence is achieved, and the program stops the estimation.

CTYPE

CTYPE=0 no termination test (default). Process goes through the full set of NBURN (SAEM or BAYES) o NITER (ITS, IMP, IMPMAP) iterations

CTYPE=1. Test for termination on objective function, thetas, and sigmas, but not on omegas.

CTYPE=2. Test for termination on objective function, thetas, sigmas, and diagonals of omegas.

CTYPE=3. Test for termination on objective function, thetas, sigmas, and all omega elements.

CINTERVAL

Every CINTERVAL iterations (default=1) is submitted to the convergence test system. If CINTERVAL is not specified, then the PRINT setting is used.

Because sequential MCMC samples are correlated, use CINTERVAL large enough to approximately "decorrelate" (such as 10-100)

CITER or CNSAMP

Number of latest PRINT or CINTERVAL iterations on which to perform a linear regression test (default=10)

If CITER=10 and CINTERVAL=100, then the most previous 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000th iterations are involved in linear regression test

CALPHA

CALPHA=0.01-0.05. Alpha error rate to use on linear regression test to assess statistical significance. The default value is 0.05, and is best to keep it at that value. The smaller the alpha, the less stringent the test, and the more likely to terminate too early.

Part E: Setting up Example 1 for Bayesian Analysis (504_bayes.ctl)

We wish to now render the full model using Bayesian analysis. As mentioned in the introduction section of this tutorial, it is strongly recommended that Bayesian analysis be performed with at least some uninformative prior information on the parameters, Particularly the OMEGA elements. The following code is added to supply uninformative priors, by adding very large prior variances $THETAPV on the THETA priors $THETAP, and by supplying low degrees of freedom $OMEGPD equal to the block dimension for the OMEGA priors $OMEGP, and a similar logic for the SIGMA priors, which can be placed immediately after the initial estimates records (control stream 504_bayes.ctl, Supplementary Material S2):

```
$SIGMA ; Initial SIGMA
  0.04
…
$PRIOR NWPRI
$THETAP (0.01 FIXED)X8
$THETAPV BLOCK(8) VALUES(100000.0,0.0) FIXED
$OMEGAP  BLOCK(2) VALUES(0.04,0.0) FIXED
```

```
$OMEGAPD (2 FIXED)
$SIGMAP BLOCK(1) (0.05 FIXED)
$SIGMAPD (1 FIXED)
```

And the BAYES estimation record in the $EST section:

```
$EST METHOD=BAYES AUTO=1 PRINT=100 NITER=10000
$COV UNCONDITIONAL MATRIX=R PRINT=E
$TABLE ID TIME IPRE CWRES ETA1 ETA2 NOPRINT ONEHEADER FILE=504_bayes.tab
```

The AUTO=1 feature with BAYES is equivalent to specifying the following options (just those different from default values are listed):

```
$EST METHOD=BAYES CTYPE=3 CINTERVAL=0
```

So, AUTO=1 turns on the convergence tester (CTYPE=3), and requests a search for the best correlation interval (CINTERVAL=0) to be used for the convergence test.

Because all fixed effects parameters are linearly MU parameterized, the efficient Gibbs sampling is utilized by NONMEM. The averages and standard deviations of the 10000 stationary distribution samples (NITER=10000) are very similar to those of FOCE, ITS, and SAEM (Table 1). See example 504_nuts in the supplementary materials S2 for using the NUTS algorithm.